Image Restoration by Matching Gradient Distributions

Image restoration by matching gradient distributions

Taeg Sang Cho, *Student Member, IEEE,* C. Lawrence Zitnick, *Member, IEEE,* Neel Joshi, *Member, IEEE*
Sing Bing Kang, *Senior Member, IEEE,* Richard Szeliski, *Fellow, IEEE,*
and William. T. Freeman, *Fellow, IEEE,*

Abstract—The restoration of a blurry or noisy image is commonly performed with a MAP estimator, which maximizes a posterior probability to reconstruct a clean image from a degraded image. A MAP estimator, when used with a sparse gradient image prior, reconstructs piecewise smooth images and typically removes textures that are important for visual realism. We present an alternative deconvolution method called *iterative distribution reweighting (IDR)* which imposes a global constraint on gradients so that a reconstructed image should have a gradient distribution similar to a reference distribution. In natural images, a reference distribution not only varies from one image to another, but also within an image depending on texture. We estimate a reference distribution directly from an input image for each texture segment. Our algorithm is able to restore rich mid-frequency textures. A large scale user study supports the conclusion that our algorithm improves the visual realism of reconstructed images compared to those of MAP estimators.

Index Terms—Non-blind deconvolution, image prior, image deblurring, image denoising

1 Introduction

Images captured with today's cameras typically contain some degree of noise and blur. In low-light situations, blur due to camera shake can ruin a photograph. If the exposure time is reduced to remove blur due to motion in the scene or camera shake, intensity and color noise may be increased beyond acceptable levels. The act of restoring an image to remove noise and blur is typically an under-constrained problem. Information lost during a lossy observation process needs to be restored with prior information about natural images to achieve visual realism. Most Bayesian image restoration algorithms reconstruct images by maximizing the posterior probability, abbreviated MAP. Reconstructed images are called the MAP estimates.

One of the most popular image priors exploits the heavy-tailed characteristics of the image's gradient distribution [7], [21], which are often parameterized using a mixture of Gaussians or a generalized Gaussian distribution. These priors favor sparse distributions of image gradients. The MAP estimator balances the observation likelihood with the gradient prior, reducing image deconvolution artifacts such as ringing and noise. The primary concern with this technique is not the prior itself, but the use of the MAP estimate. Since the MAP estimate penalizes non-zero gradients, the images often appear overly smoothed with abrupt step edges resulting in a cartoonish appearance and a loss of mid-frequency textures, Figure 1.

In this paper, we introduce an alternative image restoration strategy that is capable of reconstructing visually pleasing textures. The key idea is not to penalize gradients based on a fixed gradient prior [7], [21], but to match the reconstructed image's gradient distribution to the desired distribution [39]. That is, we attempt to find an image that lies on the manifold

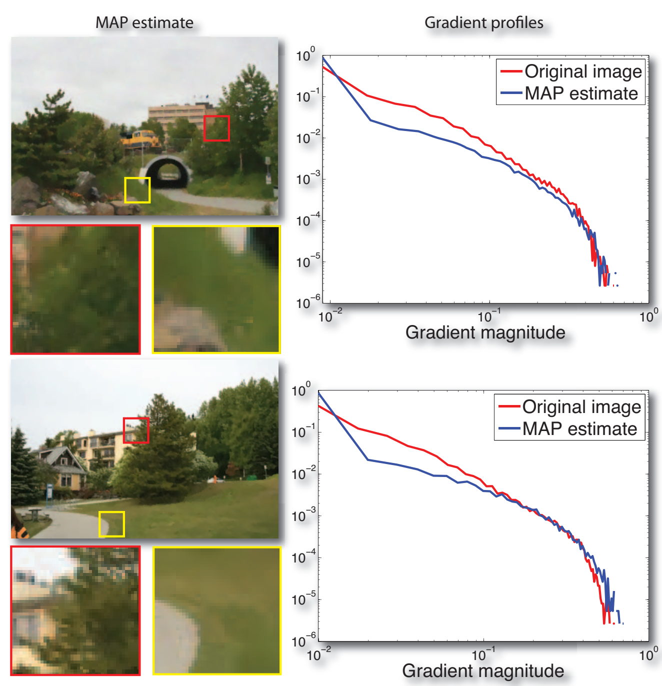


Fig. 1: *The gradient distribution of images reconstructed using the MAP estimator can be quite different from that of the original images. We present a method that matches the reconstructed image's gradient distribution to that of the desired gradient distribution (in this case, that of the original image) to hallucinate visually pleasing textures.*

• *T. S. Cho and W. T. Freeman are with Computer Science and Artificial Intelligence Lab (CSAIL), Massachusetts Institute of Technology, Cambridge, MA 02139.*

• *C. L. Zitnick, N. Joshi, S. B. Kang and R. Szeliski are with Microsoft Research, Redmond, WA 98004.*

Digital Object Indentifier 10.1109/TPAMI.2011.166 0162-8828/11/$26.00 © 2011 IEEE

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

of solutions with the desired gradient distribution, which maximizes the observation likelihood. We propose two approaches. The first penalizes the gradients based on the KL divergence between the empirical and desired distributions. Unfortunately, this approach may not converge or may find solutions with gradient distributions that vary significantly from the desired distribution. Our second approach overcomes limitations of the first approach by defining a cumulative penalty function that gradually pushes the parameterized empirical distribution towards the desired distribution. The result is an image with a gradient distribution that closely matches that of the desired distribution.

A critical problem in our approach is determining the desired gradient distribution. To do this we borrow a heuristic from Cho *et. al.* [4] that takes advantage of the fact that many textures are scale invariant. A desired distribution is computed using a downsampled version of the image over a set of segments. We demonstrate our results on several image sets with both noise and blur. Since our approach synthesizes textures or gradients to match the desired distribution, the peak signal-to-noise ratio (PSNR) and gray-scale SSIM [37] may be below other techniques. However, the results are generally more visually pleasing. We validate these claims using a user study comparing our technique to those reconstructed using the MAP estimator.

2 Related work

2.1 Image denoising

Numerous approaches to image denoising have been proposed in the literature. Early methods include decomposing the image into a set of wavelets. Low amplitude wavelet values are simply suppressed to remove noise in a method call coring [30], [35]. Other techniques include anisotropic diffusion [26] and bilateral filtering [36]. Both of these techniques remove noise by only blurring neighboring pixels with similar intensities, resulting in edges remaining sharp. The FRAME model [41] showed Markov Random Field image priors can be learned from image data to perform image reconstruction. Recently, the Field of Experts approach [31] proposed a technique to learn generic and expressive image priors for traditional MRF techniques to boost the performance of denoising and other reconstruction tasks.

The use of multiple images has also been proposed in the literature to remove noise. Petschnigg *et. al.* [27] and Eisemann *et. al.* [6] proposed combining a flash and non-flash image to produce reduced noise and naturally colored images. Bennett *et. al.* [1] use multiple frames in a video to denoise, while Joshi and Cohen [14] combined hundreds of still images to create a single sharp and denoised image. In this paper, we only address the tasks of denoising and deblurring from a single image.

2.2 Image deblurring

Blind image deconvolution is the combination of two problems: estimating the blur kernel or PSF, and image deconvolution. A survey of early work in these areas can be found in Kundur and Hatzinakos [18]. Recently, several works have used gradient priors to solve for the blur kernel and to aid in deconvolution [7], [16], [21]. We discuss these in more detail in the next section. Joshi *et. al.* [15] constrained the computation of the blur kernel resulting from camera shake using additional hardware. A coded aperture [21] or fluttered shutter [29] may also be used to help in the estimation of the blur kernel or in deconvolution. A pair of images with high noise (fast exposure) and camera shake (long exposure) was used by Yuan *et. al.* [40] to aid in constraining deconvolution. Approaches by Whyte *et. al.* [38] and Gupta *et. al.* [11] attempt to perform blind image deconvolution with spatially variant blur kernels, unlike most previous techniques that assume spatially invariant kernels. In our work, we assume the blur kernel, either spatially variant or invariant, is known or computed using another method. We only address the problem of image deconvolution.

2.3 Gradient priors

The Wiener filter [10] is a popular image reconstruction method with a closed form solution. The Wiener filter is a MAP estimator with a Gaussian prior on image gradients, which tends to blur edges and causes ringing around edges because those image gradients are not consistent with a Gaussian distribution.

Bouman and Sauer [2], Chan and Wong [3], and more recently Fergus *et. al.* [7] and Levin *et. al.* [21], use a heavy-tailed gradient prior such as a generalized Gaussian distribution [2], [21], a total variation [3], or a mixture of Gaussians [7]. MAP estimators using sparse gradient priors preserve sharp edges while suppressing ringing and noise. However, they also tend to remove mid-frequency textures, which causes a mismatch between the reconstructed image's gradient distribution and that of the original image.

2.4 Matching gradient distributions

Matching gradient distributions has been addressed in the texture synthesis literature. Heeger and Bergen [13] synthesize textures by matching wavelet sub-band histograms to those of the desired texture. Portilla and Simoncelli [28] match joint statistics of wavelet coefficients to synthesize homogeneous textures. Kopf *et. al.* [17] introduce a non-homogeneous texture synthesis technique by matching histograms of texels (or elements of textures).

Matching gradient distributions in image restoration is not entirely new. Li and Adelson [22] introduce a two-step image restoration algorithm that first reconstructs an image using an exemplar-based technique similar to Freeman *et. al.* [9], and warps the reconstructed image's gradient distribution to a reference gradient distribution using Heeger and Bergen's method [13].

A similarly motivated technique to ours is proposed by Woodford *et. al.* [39]. They use a MAP estimation framework called a marginal probability field (MPF) that matches a histogram of low-level features, such as gradients or texels, for computer vision tasks including denoising. While both Woodford *et. al.* and our techniques use a global penalty term to fit the global distribution, MPF requires that one bins features to form a discrete histogram. This may lead to artifacts with small gradients. Our distribution matching method by-passes this binning process using parameterized continuous functions. Also, Woodford *et. al.* [39] use an image prior estimated from a database of images and use the same global prior to reconstruct images with different textures. In contrast, we estimate the image prior directly from the degraded image for each textured region. Schmidt *et. al.* [34] match the gradient distribution through sampling, which may be computationally expensive in practice. As with Woodford *et. al.* [39], Schmidt *et. al.* also use a single global prior to reconstruct images with different textures, which causes noisy renditions in smooth regions. HaCohen *et. al.* [12] explicitly integrate texture synthesis to image restoration, specifically for an image up-sampling problem. To restore textures, they segment a degraded image and replace each texture segment with textures in a database of images.

3 Characteristics of MAP estimators

In this section, we illustrate why MAP estimators with a sparse prior recover unrealistic, piecewise smooth renditions as illustrated in Figure 1. Let B be a degraded image, k be a blur kernel, $\otimes$ be a convolution operator, and I be a latent image. A MAP estimator corresponding to a linear image observation model and a gradient image prior solves the following regularized problem:

$$\hat{I} = \underset{I}{\operatorname{argmin}} \left\{ \frac{\|B - k \otimes I\|^2}{2\eta^2} + w \sum_m \rho(\nabla_m I) \right\}, \quad (1)$$

where η^2 is an observation noise variance, m indexes gradient filters, and ρ is a robust function that favors sparse gradients. We parameterize the gradient distribution using a generalized Gaussian distribution. In this case, $\rho(\nabla I) = -\ln(p(\nabla I; \gamma, \lambda))$, where the prior $p(\nabla I; \gamma, \lambda)$ is given as follows:

$$p(\nabla I; \gamma, \lambda) = \frac{\gamma \lambda^{(\frac{1}{\gamma})}}{2\Gamma(\frac{1}{\gamma})} \exp(-\lambda |\nabla I|^\gamma). \quad (2)$$

Γ is a Gamma function and shape parameters γ, λ determine the shape of the distribution. In most MAP-based image reconstruction algorithms, gradients are assumed to be independent for computational efficiency: $p(\nabla I; \gamma, \lambda) = \frac{1}{Z} \prod_{i=1}^N p(\nabla I_i; \gamma, \lambda)$, where i is a pixel index, Z is a partition function, and N is the total number of pixels in an image.

A MAP estimator balances two competing forces: the reconstructed image $\hat{I}$ should satisfy the observation model while conforming to the image prior. Counter-intuitively, the image prior term, assuming independence among gradients, *always* favors a flat image to any other image, even a natural image. Therefore, the more the MAP estimator relies on the image prior term, which is often the case when the image degradation is severe, the more the reconstructed image becomes piecewise smooth.

One way to explain this property is that the independence among local gradients fails to capture the global statistics of gradients for the whole image. The image prior tells us that gradients in a natural image *collectively* exhibit a sparse gradient profile, whereas the independence assumption of gradients forces us to minimize each gradient *independently*, always favoring a flat image. Nikolova [25] provides a theoretic treatment of MAP estimators in general to show its deficiency.

We could remove the independence assumption and impose a joint prior on all gradients, but this approach is computationally expensive. This paper introduces an alternative method to impose a global constraint on gradients – that a reconstructed image should have a gradient distribution similar to a reference distribution.

4 Image reconstruction

In this section, we develop an image reconstruction algorithm that minimizes the KL divergence between the reconstructed image's gradient distribution and its reference distribution. This distance penalty plays the role of a global image prior that steers the solution away from piecewise smooth images.

Let $q_E(\nabla I)$ be an empirical gradient distribution of an image I, and q_D be a reference or desired distribution. We measure the distance between distributions q_E and q_D using the Kullback-Leibler (KL) divergence:

$$KL(q_E || q_D) = \int_x q_E(x) \ln\left(\frac{q_E(x)}{q_D(x)}\right) dx. \quad (3)$$

An empirical distribution q_E is parameterized using a generalized Gaussian distribution $p(\nabla I; \gamma, \lambda)$ (Eq. 2). Given gradient samples, ∇I_i, where i indexes samples, we estimate the shape parameters γ_E, λ_E of an empirical gradient distribution q_E by maximizing the log-likelihood:

$$[\gamma_E, \lambda_E] = \underset{\gamma, \lambda}{\operatorname{argmin}} \left\{ -\sum_{i=1}^N \frac{1}{N} \ln(p(\nabla I_i; \gamma, \lambda)) \right\}. \quad (4)$$

This is equivalent to minimizing the KL divergence between gradient samples ∇I and a generalized Gaussian distribution. We use the Nelder-Mead optimization method [19] to solve Eq. 4.

4.1 Penalizing the KL divergence directly

To motivate our algorithm in Section 4.2, we first introduce a method that penalizes the KL divergence between an empirical gradient distribution q_E and a reference distribution q_D. We

Algorithm 1 MAP with KL penalty

% Initial image estimate to start iterative minimization
$\hat{I}^0 = \text{argmin}_I \left\{ \frac{\|B-k\otimes I\|^2}{2\eta^2} + w_1\lambda_D|\nabla I|^{\gamma_D} \right\}$
Update ${q_E}^0$ using Eq. 4
% Iterative minimization
for l = 1 ... 10 **do**
  % KL distance penalty term update
  $\rho_G^l(\nabla I) = \frac{1}{N}\ln\left(\frac{{q_E}^{(l-1)}(\nabla I)}{q_D(\nabla I)}\right)$
  % Image reconstruction
  $\hat{I}^l = \text{argmin}_I \left\{ \frac{\|B-k\otimes I\|^2}{2\eta^2} + w_1\lambda_D|\nabla I|^{\gamma_D} + w_2\rho_G^l(\nabla I) \right\}$
  Update ${q_E}^l$ using Eq. 4
end for
$\hat{I} = \hat{I}^{10}$

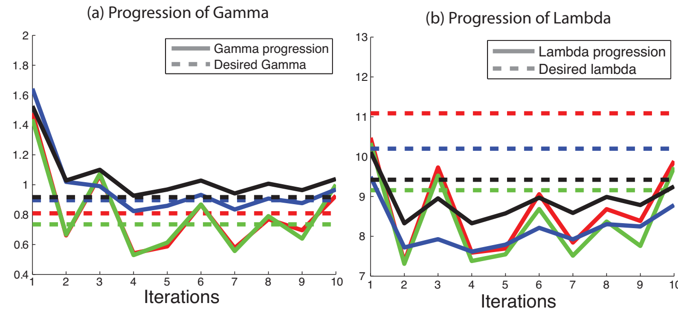


Fig. 3: *We illustrate the operation of Algorithm 1 in terms of the γ_E, λ_E progressions. Different colors correspond to different gradient filters. Oftentimes, Algorithm 1 does not converge to a stable point, but oscillates around the desired solution.*

show that the performance of this algorithm is sensitive to the parameter setting and that the algorithm may not always converge. In Section 4.2, we extend this algorithm to a more stable approach called Iterative Distribution Reweighting (IDR) for which the found empirical distribution is closer to q_D.

We can penalize the KL divergence between q_E and q_D by adding a term to the MAP estimator in Eq. 1

$$\hat{I} = \underset{I}{\text{argmin}} \left\{ \frac{\|B-k\otimes I\|^2}{2\eta^2} + w_1\lambda_D|\nabla I|^{\gamma_D} + w_2 KL(q_E||q_D) \right\}, \quad (5)$$

where w_2 determines how much to penalize the KL divergence.[1] It's hard to directly solve Eq. 5 because the KL divergence is a non-linear function of a latent image I. Therefore we solve Eq. 5 iteratively.

Using the set ∇I as a non-parametric approximation of q_E and Eq. 3, we estimate $KL(q_E||q_D)$ using

$$KL(q_E||q_D) \approx \sum_i^N \rho_G(\nabla I_i) = \sum_i^N \left\{ \frac{1}{N}\ln\left(\frac{q_E(\nabla I_i)}{q_D(\nabla I_i)}\right) \right\}, \quad (6)$$

where $\rho_G(\nabla I_i)$ is the energy associated with a KL divergence for each gradient sample ∇I_i.

Algorithm 1 shown using pseudocode, iteratively computes the values of $\rho_G(\nabla I_i)$ using the previous iteration's empirical distribution ${q_E}^{(l-1)}$, followed by solving Eq. 5. The accuracy of our approximation of $KL(q_E||q_D)$ is dependent on two factors. The first is the number of samples in ∇I. As we discuss later in Section 4.3 we may assume a significant number of samples, since the value of Eq. 6 is computed over large segments in the image. Second, the parametrization of q_E is computed from the previous iteration's samples. As a result, the approximation becomes more accurate as the approach converges.

Using $\rho_G(\nabla I)$, we can describe Algorithm 1 qualitatively as follows: if q_E has more gradients of a certain magnitude than q_D, ρ_G penalizes those gradients *more*; if q_E has fewer gradients of a certain magnitude than q_D, they receive *less* penalty. Therefore, the approach favors distributions q_E close to q_D. Figure 2 illustrates the procedure. The full derivation of the algorithm details is available in the supplemental material.

4.1.1 Algorithm analysis

To provide some intuition for the behavior of Algorithm 1, consider the case when q_E approaches q_D. The cost function ρ_G will approach zero. The result is a loss of influence for the cost related to the KL divergence, and q_E may not fully converge to q_D. q_E can be forced arbitrarily close to q_D by increasing the weight w_2 and reducing the influence of the other terms. Unfortunately, when w_2 is large, the algorithm oscillates around the desired solution (Figure 3). Even if under-relaxation techniques are used to reduce oscillations, q_E may be significantly different from q_D for reasonable values of w_2. If w_2 is too large, the linearized system (in supplemental material, (11)) becomes indefinite, in which case the minimum residual method [33] cannot be used to solve the linearized system. To mitigate the reliability issue and to damp possible oscillations around the desired solution, we develop an iterative distribution reweighting algorithm.

4.2 The iterative distribution reweighting (IDR)

In this section, we propose a second approach called Iterative Distribution Reweighting (IDR) that solves many of the shortcomings of Algorithm 1. Previously, we minimized a global energy function that only penalized empirical distributions that diverged from q_D. As discussed in Section 4.1.1, this approach may not converge, or upon convergence the found gradient distribution may vary significantly from q_D. Our second approach can be interpreted as minimizing the data cost function from Eq. 1, while actively pushing the parameterized empirical distribution q_E towards our reference distribution q_D,

$$\hat{I} = \underset{I}{\text{argmin}} \left\{ \frac{\|B-k\otimes I\|^2}{2\eta^2} \right\}, \quad (7)$$

1. In Eq. 5, we have replaced the summation over multiple filters in Eq. 1, i.e. $\sum_m \lambda_m|\nabla_m I|^{\gamma_m}$, with a single derivative filter to reduce clutter, but the derivation can easily be generalized to using multiple derivative filters. We use four derivative filters in this work: x, y derivative filters and x-y, and y-x diagonal derivative filters.

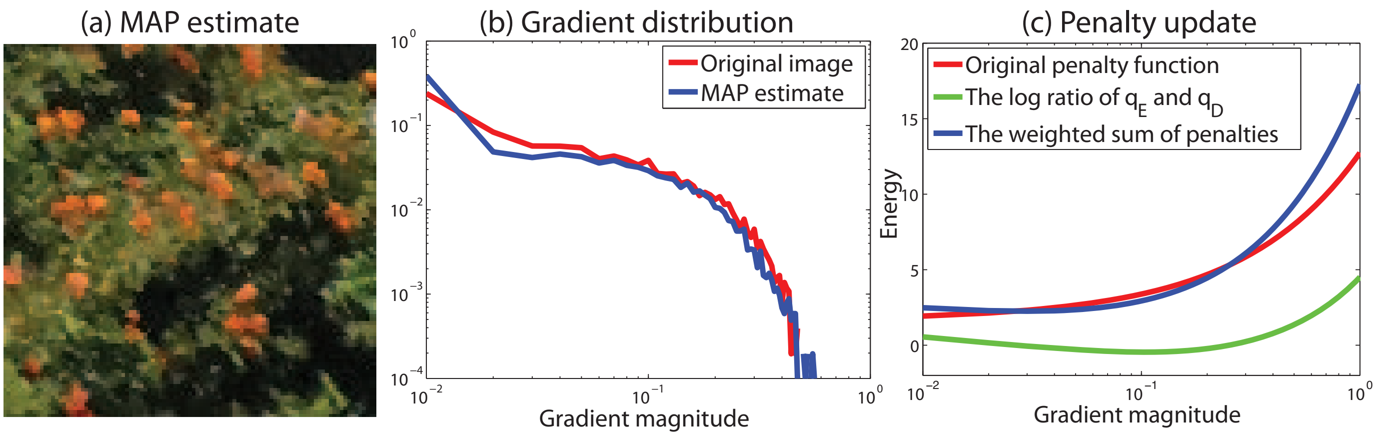


Fig. 2: *This figure illustrates Algorithm 1. Suppose we deconvolve a degraded image using a MAP estimator. (b) shows that the x-gradient distribution of the MAP estimate in (a) does not match that of the original image. (c) Our algorithm adds the log ratio of q_E and q_D to the original penalty (i.e., $\lambda_D|\nabla I|^{\gamma_D}$) such that the weighted sum of the two penalty terms encourages a better distribution match in the following iteration. q_D is set to the ground truth distribution.*

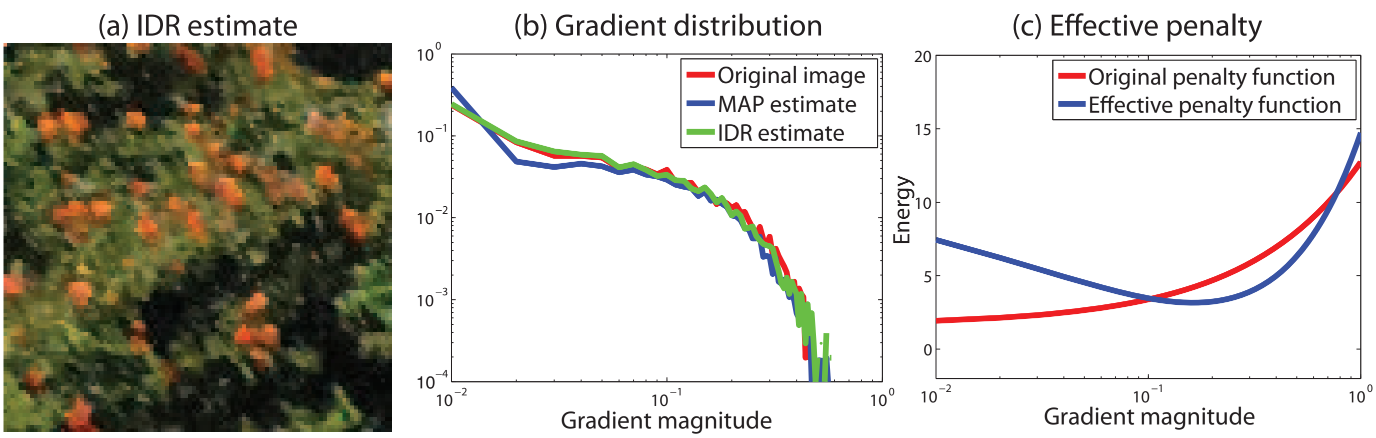


Fig. 4: *The IDR deconvolution result. (a) shows the deconvolved image using IDR, and (b) compares the gradient distribution of images reconstructed using the MAP estimator and IDR. (c) The effective penalty after convergence (i.e. $w_1\lambda_D|\nabla I|^{\gamma_D} + w_2\sum_{l=1}^{10}\frac{1}{N}\ln\left(\frac{q_E{}^l(\nabla I)}{q_D(\nabla I)}\right)$) penalizes gradients with small and large magnitude more than gradients with moderate magnitude. q_D is set to the ground truth distribution.*

$$s.t.\ q_E = q_D.$$

That is, our goal is to find a solution that lies on the manifold of solutions defined by $q_E = q_D$ that minimizes Eq. 7. In this paper, we do not claim to find the global minimum along the manifold, but in practice we find our heuristic to provide solutions that have a low energy with $q_E \approx q_D$.

While conceptually quite different from Algorithm 1, the approaches are similar in implementation. As in the KL divergence term of Algorithm 1, we add an additional cost function to Eq. 7 using the ratio of the distributions q_E and q_D. However, instead of penalizing the KL divergence between q_E and q_D directly, we propose a new cumulative cost function $\hat{\rho}_G$. During each iteration, we update $\hat{\rho}_G$ to push q_E closer to q_D by examining the parameterized empirical distribution from the previous iteration. For instance, if the empirical probability of a set of gradients is too high relative to q_D in the current iteration, their penalty is increased in the next iteration. Our new cost function $\hat{\rho}_G^l$ is

$$\hat{\rho}_G^l(\nabla I) = \hat{\rho}_G^{(l-1)}(\nabla I) + w_2\frac{1}{N}\ln\left(\frac{q_E{}^{(l-1)}(\nabla I)}{q_D(\nabla I)}\right), \quad (8)$$

where

$$\hat{\rho}_G^0(\nabla I) = w_1\lambda_D|\nabla I|^{\gamma_D}. \quad (9)$$

The first term of Eq. 8 is the cost function from the previous iteration. The second term updates the cost function using the ratio between q_D and the parameterized gradient distribution resulting from the use of $\hat{\rho}_G^{(l-1)}$. We initialize $\hat{\rho}_G^0$ using the gradient prior from Eq. 1 to bias at the outset results with sparse gradients. In practice λ_D and γ_D my be set using the parameters of the reference distribution, or simply set to some default values. As discussed in Section 4.3, we kept them fixed to default values for use in estimating q_D. Combining Equation Eq. 7 with our new cost function $\hat{\rho}_G$, our new approach

Algorithm 2 The iterative distribution reweighting (IDR)

% Initial image estimate to start iterative minimization

$\hat{I}^0 = \text{argmin}_I \left\{ \frac{\|B-k\otimes I\|^2}{2\eta^2} + w_1 \lambda_D |\nabla I|^{\gamma_D} \right\}$

Update ${q_E}^0$ using Eq. 4

% Iterative minimization

for l = 1 ... 10 **do**

  % Accumulating the KL divergence

  $\hat{\rho}_G^l(\nabla I) = \hat{\rho}_G^{(l-1)}(\nabla I) + w_2 \frac{1}{N} \ln\left(\frac{{q_E}^{(l-1)}(\nabla I)}{q_D(\nabla I)}\right)$

  % Image reconstruction

  $\hat{I}^l = \text{argmin}_I \left\{ \frac{\|B-k\otimes I\|^2}{2\eta^2} + \rho_G^l(\nabla I) \right\}$

  Update ${q_E}^l$ using Eq. 4

end for

$\hat{I} = \hat{I}^{10}$

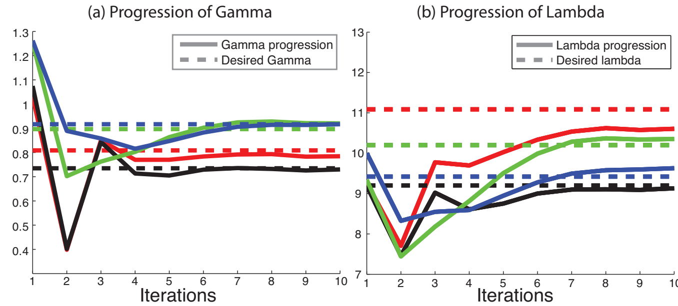


Fig. 5: *This figure shows how the γ_E, λ_E progress from one iteration to the next. Different colors correspond to different gradient filters. We observe that the algorithm converges to a stable point in about 8 iterations.*

iteratively solves

$$\hat{I} = \underset{I}{\text{argmin}} \left\{ \frac{\|B - k \otimes I\|^2}{2\eta^2} + \hat{\rho}_G(\nabla I) \right\}, \tag{10}$$

as shown in pseudocode by Algorithm 2. IDR iteratively adjusts the penalty function $\hat{\rho}_G$ by the ratio of distributions q_E and q_D using a formulation similar to the previous approach using KL divergence Eq. 6, thus the name *iterative distribution reweighting (IDR)*. The detailed derivations in the supplemental material, Section 3, can be easily modified for use with Algorithm 2.

Examining Eq. 8, if the parameterized empirical distribution q_E is equal to q_D, $\hat{\rho}_G^l$ is equal to the cost function from the previous iteration, $\hat{\rho}_G^{l-1}$. As a result, the desired solution $q_E = q_D$ is a stable point for IDR[2]. It is worth noting that when $q_E = q_D$, $\hat{\rho}_G$ will not be equal to the sparse gradient prior, as occurs for the gradient priors in Algorithm 1 since $\rho_G = 0$. Consequently, Algorithm 2 can converge to solutions with q_E arbitrarily close to q_D for various values of w_2. The value of w_2 may also be interpreted differently for both algorithms. In Algorithm 1, w_2 controls the strength of the bias of q_E towards q_D, where w_2 controls the rate q_E converges to q_D in Algorithm 2. That is, even for small values of w_2, Algorithm 2 typically converges to $q_E \approx q_D$.

We illustrate the operation of IDR in Figure 4, and show how γ_E, λ_E changes from one iteration to the next in Figure 5. Observe that γ_E, λ_E no longer oscillates as in Figure 3. In Figure 4, we show the original penalty function and its value after convergence. Note it is not equal to the sparse gradient prior and significantly different from the penalty function found by Algorithm 1, Figure 2.

In Figure 6, we test IDR for deblurring a single texture, assuming that the reference distribution q_D is known a priori. We synthetically blur the texture using the blur kernel shown in Figure 8 and add 5% Gaussian noise to the blurred image. We deblur the image using a MAP estimator and using IDR, and compare the reconstructions. For all examples in this paper, we use $w_1 = 0.025, w_2 = 0.0025$. We observe that the gradient distribution of the IDR estimate matches the reference distribution better than that of the MAP estimate, and visually, the texture of the IDR estimate better matches the original image's texture. Although visually superior, the peak signal-to-noise ratio (PSNR) and gray-scale SSIM [37] of the IDR estimate are lower than those of the MAP estimate. This occurs because IDR may not place the gradients at exactly the right position. Degraded images do not strongly constrain the position of gradients, in which case our algorithm disperses gradients to match the gradient distribution, resulting in lower PSNR and SSIM measures.

2. This statement does not mean that the algorithm will converge only if $q_E = q_D$; the algorithm can converge to a local minimum.

4.2.1 Algorithm analysis

IDR matches a *parametrized* gradient distribution q_E, and therefore the algorithm is inherently limited by the accuracy of the fit. The behavior of IDR is relatively insensitive to the weighting term w_2, since w_2 no longer controls how close q_E is to q_D, but the rate at which q_E approaches q_D. Similarly to Algorithm 1, a large w_2 can destabilize the minimum residual algorithm [33] that solves the linearized system in Supplemental material, (11).

In most cases, IDR reliably reconstructs images with the reference gradient distribution. However, there are cases in which the algorithm settles at a local minimum that does not correspond to the desired texture. This usually occurs when the support of the derivative filters is large and when we use many derivative filters to regularize the image. For instance, suppose we want to match the gradient histogram of a 3×3 filter. The algorithm needs to update 9 pixels to change the filter response at the center pixel, but updating 9 pixels also affects filter the responses of 8 neighboring pixels. Having to match multiple gradient distributions at the same time increases the complexity and reduces the likelihood of convergence. To control the complexity, we match four two-tap derivative filters. Adapting derivative filters to local image structures using steerable filters [4], [8], [32] may further improve the rendition of oriented textures, but it is not considered in this work.

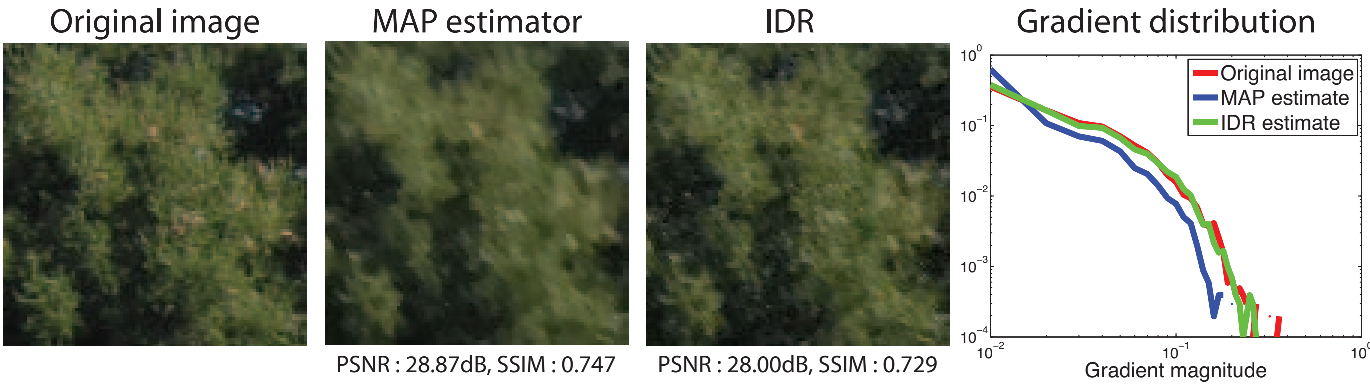


Fig. 6: *We compare the deblurring performance of a MAP estimator and IDR. IDR reconstructs visually more pleasing mid-frequency textures compared to a MAP estimator.*

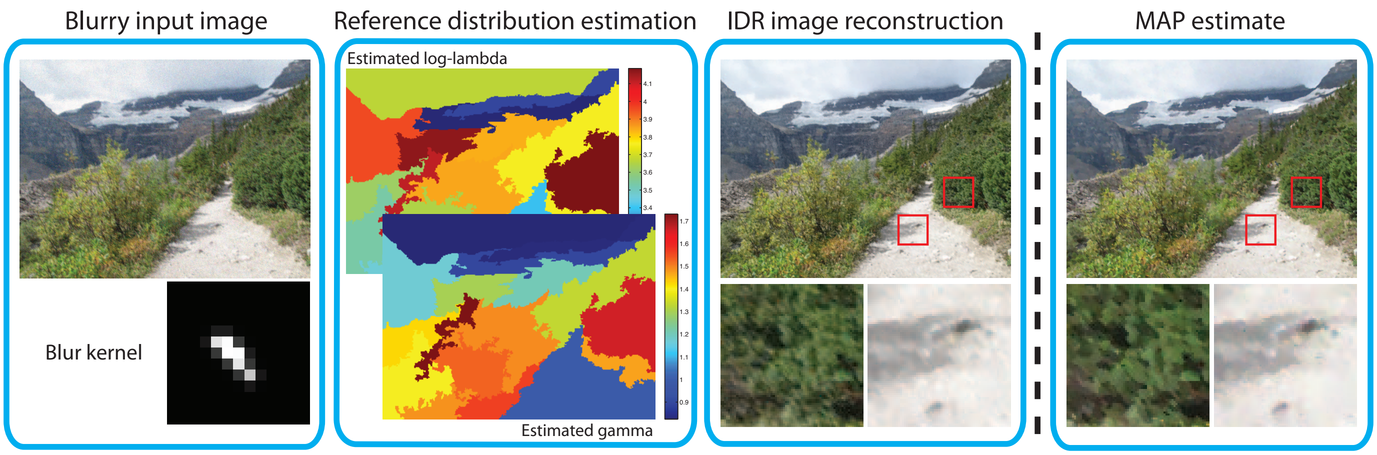


Fig. 7: *For an image with spatially varying texture, our algorithm segments the image into regions of homogeneous texture and matches the gradient distribution in each segment independently. Compared to MAP estimators, our algorithm reconstructs visually more pleasing textures.*

4.3 Reference distribution q_D estimation

We parameterize a reference distribution q_D using a generalized Gaussian distribution. Unfortunately, one often does not know *a priori* what q_D should be. Previous work estimates q_D from a database of natural images [7], [39] or hand-picks q_D through trial and error [21]. We adopt the image prior estimation technique introduced in Cho *et. al.* [4] to estimate q_D directly from a degraded image, as we will now describe.

It is known that many textures are scale invariant due to the fractal properties of textures and piecewise smooth properties of surfaces [20], [24]. That is, the gradient profiles are roughly equal across scales, whereas the affect of deconvolution noise tends to be scale variant. Cho *et. al.* [4] propose deconvolving an image, followed by downsampling. The downsampled image is then used to estimate the gradient distribution. The result is the scale invariant gradient distribution is maintained, while the noise introduced by deconvolution is reduced during downsampling. This approach will result in incorrect distributions for textures that are not scale invariant, such as brick textures, but produces reasonable results for many real-world textures.

When deconvolving the degraded image B we use a MAP estimator (Eq. 1) with a hand-picked image prior, tuned to restore different textures reasonably well at the expense of a slightly noisy image reconstruction (i.e., a relatively small gradient penalty). In this paper, we set the parameters of the image prior as $[\gamma = 0.8, \lambda = 4, w_1 = 0.01]$ for all images. We fit gradients from the down-sampled image to a generalized Gaussian distribution, as in Eq. 4, to estimate the reference distribution q_D. While fine details can be lost through down-sampling, empirically, the estimated reference distribution q_D is accurate enough for our purpose.

Our image reconstruction algorithm assumes that the texture is homogeneous (i.e., a single q_D). In the presence of multiple textures within an image, we segment the image and estimate separate reference distributions q_D for each segment: we use the EDISON segmentation algorithm [5] to segment an image into about 20 regions. Figure 7 illustrates the image deconvolution process for spatially varying textures. Unlike Cho *et. al.* [4] we cannot use a per-pixel gradient prior, since we need a large area of support to compute a parameterized empirical distribution q_E in Eq. 8. However, Cho *et. al.* [4] use the standard MAP estimate, which typically does not result in

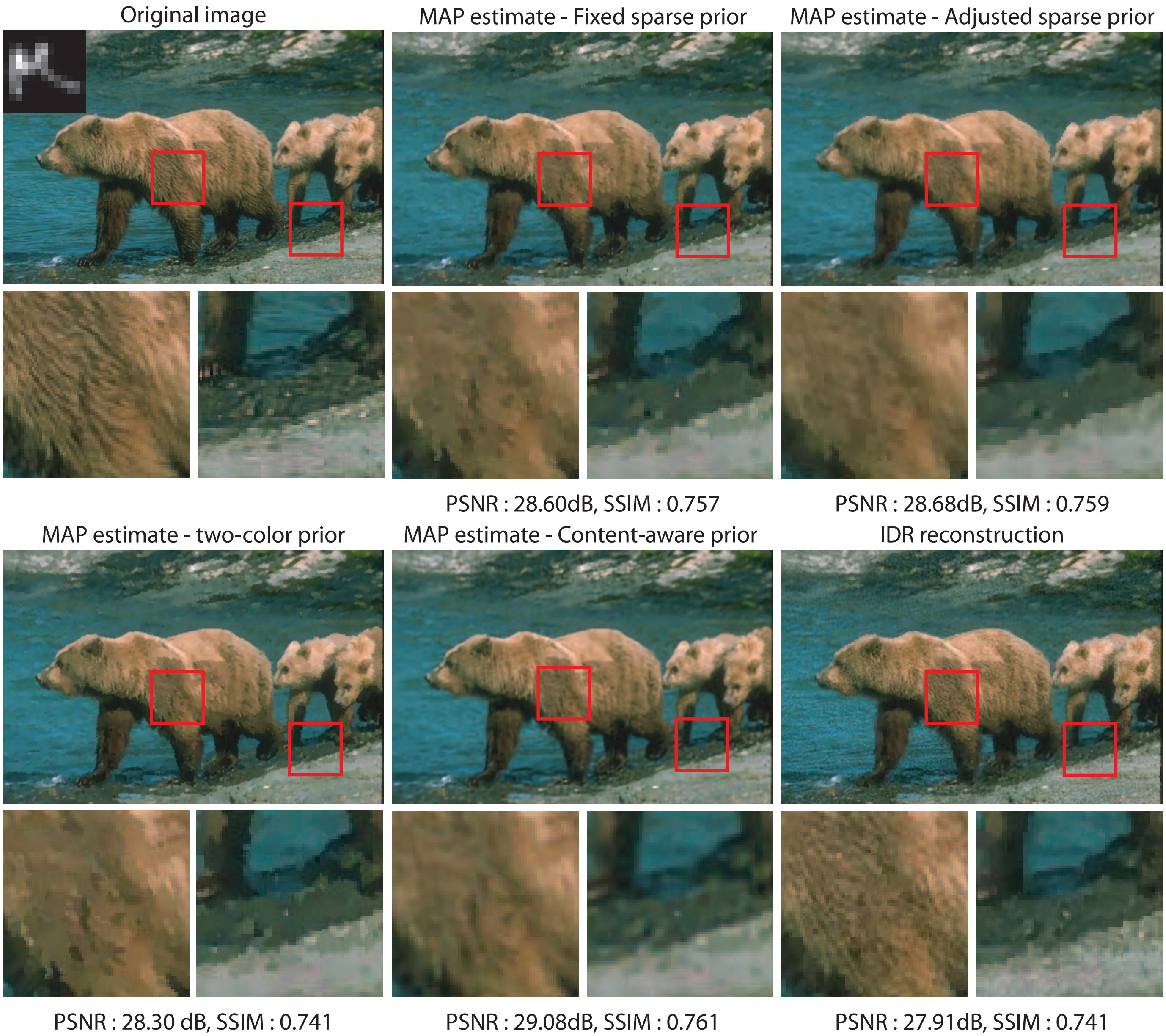


Fig. 8: *We compare the performance of IDR against four other competing methods: (i) a MAP estimator with a sparse gradient prior [21], (ii) a MAP estimator with a sparse prior adapted to each segment, (iii) a MAP estimator with a two-color prior [16], (iv) a MAP estimator with a content-aware image prior. The red box indicate the cropped regions. Although the PSNR and the SSIM of our results are often lower than those using MAP estimators, IDR restores more visually pleasing textures (see bear furs).*

images that contain the desired distribution.

5 Experiments

5.1 Deconvolution experiments

We synthetically blur sharp images with the blur kernel shown in Figure 8, add 2% noise, and deconvolve them using competing methods. We compare the performance of IDR against four other competing methods: (i) a MAP estimator with a sparse gradient prior [21], (ii) a MAP estimator with a sparse prior adapted to each segment (iii) a MAP estimator with a two-color prior [16] (iv) a MAP estimator with a content-aware image prior [4]. We blur a sharp image using the kernel shown on the right, add 2% noise to it, and restore images using the competing methods. Figure 8 shows experimental results. As mentioned in Section 4.2, IDR does not perform the best in terms of PSNR / SSIM. Nevertheless, IDR reconstructs mid-frequency textures better, for instance fur details. Another interesting observation is that the content-aware image prior performs better, in terms of PSNR/SSIM, than simply adjusting the image prior to each segment's texture. By using the segment-adjusted image prior, we observe segmentation boundaries that are visually disturbing. Another

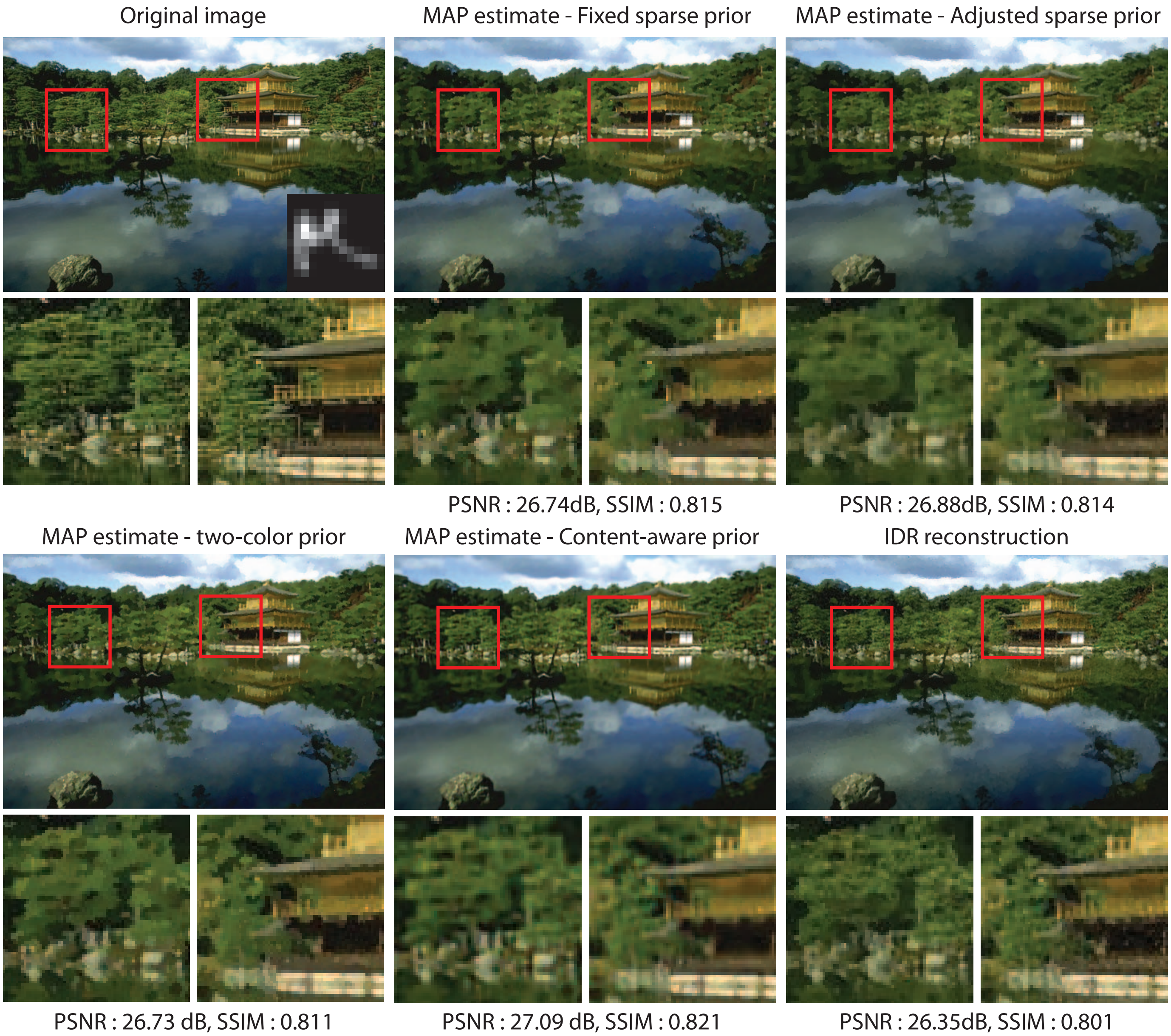


Fig. 9: *We compare the performance of IDR against four other competing methods. As in Figure 8, IDR's PSNR/SSIM are lower than those of MAP estimators, but IDR restores visually more pleasing textures.*

set of comparisons is shown in Figure 9.

In Figure 10, we compare the denoising performance of IDR to that of a marginal probability field (MPF) by Woodford *et. al.* [39] at two noise levels (their implementation only handles grayscale, square images). Using MPF for denoising has two drawbacks. First, MPF quantizes intensity levels and gradient magnitudes to reduce computation. MPF quantizes 256 (8-bit) intensity levels to 64 intensity levels (6-bit), and it bins 256 (8-bit) gradient magnitudes to 11 slots. These quantizations can accentuate spotty noise in reconstructed images. IDR adopts a continuous optimization scheme that does not require any histogram binning or intensity quantization, therefore it does not suffer from quantization noise. Second, Woodford *et. al.* [39] estimate the reference gradient distribution from a database of images, and use the *same* prior to denoise different images. This can be problematic because different images have different reference distributions q_D, but MPF would enforce the same gradient profile on them. Also, MPF does not adapt the image prior to the underlying texture, treating different textures the same way. Therefore, MPF distributes gradients uniformly across the image, even in smooth regions, which can be visually disturbing. IDR addresses these issues by estimating a reference distribution q_D from an input image and by adapting q_D to spatially varying texture.

At a high degradation level, such as a noise level of 31.4%, our reference distribution estimation algorithm can be unstable. In Figure 10(a), our q_D estimation algorithm returns a distribution that has more “large” derivatives and fewer “small” derivatives

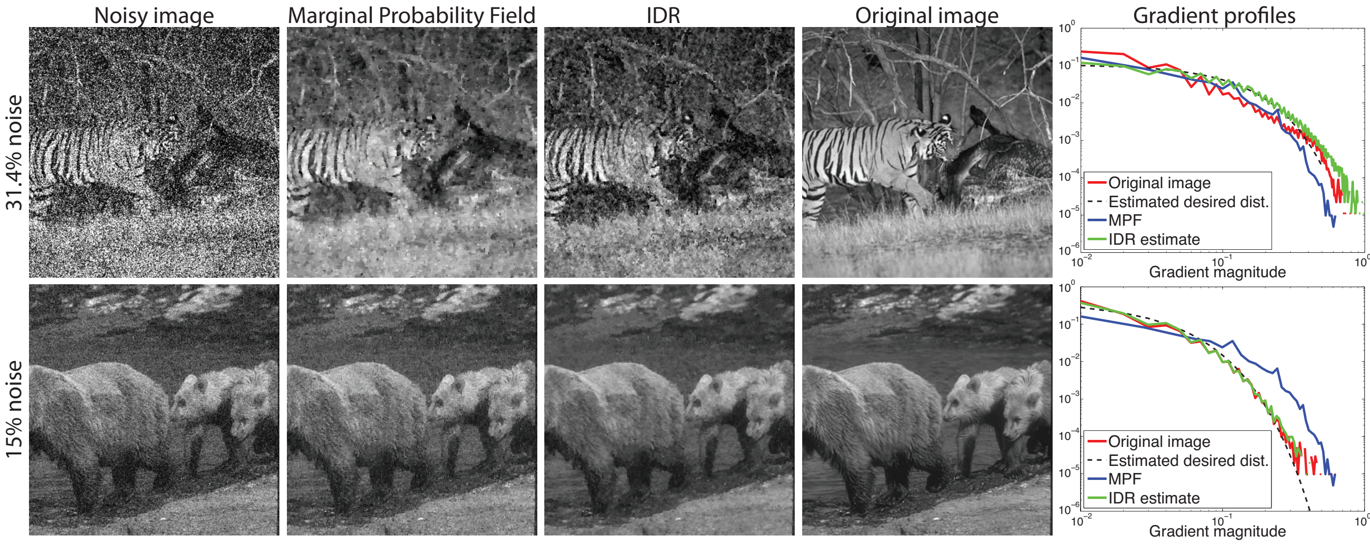


Fig. 10: *Comparing the denoising performance of IDR to the marginal probability field (MPF) [39]. IDR generates a better rendition of the spatially variant texture.*

(dotted line in Figure 10), which manifests itself as a noisy IDR reconstruction. In contrast, MPF restores a plausible image, but this is somewhat coincidental in that the reference distribution that MPF imposes is quite similar to that of the original image.

At a more reasonable degradation level (15% noise), shown in Figure 10(b), our algorithm estimates a reference distribution that is very similar to that of the original image. Given a more accurate reference distribution, IDR restores a visually pleasing image. On the other hand, MPF restores a noisy rendition because the reference distribution is quite different from that of the original image. Also note that the gradient distribution of the restored image in Figure 10(b) is very similar to that of the restored image in Figure 10(a), which illustrates our concern that using a single image prior for different images would degrade the image quality.

In this work, we estimate the reference distribution q_D assuming that the underlying texture is scale-invariant. Although this assumption holds for fractal textures, it does not strictly hold for other types of textures with a characteristic scale, such as fabric clothes, ceramics, or construction materials. The IDR algorithm is decoupled from the reference distribution estimation algorithm. Therefore, if an improved reference distribution estimation algorithm is available, the improved algorithm can be used in place of the current distribution algorithm without impacting the IDR algorithm itself.

Segmenting images to regions and deconvolving each region separately may generate artificial texture boundaries, as in Figure 11. While this rarely occurs, we could mitigate these artifacts using a texture-based segmentation algorithm rather than EDISON [5], which is a color-based segmentation algorithm.

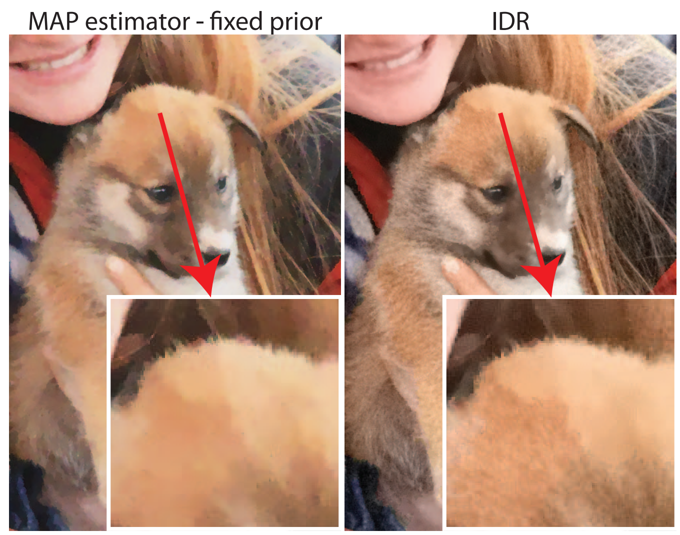


Fig. 11: *We could observe an artificial boundary when the estimated prior is different in adjacent segments that have similar textures. While this rarely occurs, we could remove such artifacts using a texture segmentation algorithm instead of a color-based segmentation algorithm.*

5.2 User study

IDR generates images with rich texture but with lower PSNR/SSIM than MAP estimates. To test our impression that images reconstructed by IDR are more visually pleasing, we performed a user study on Amazon Mechanical Turk.

We considered seven image degradation scenarios: noisy observations with 5%, 10%, 15% noise, blurry observations with a small blur and 2%, 5%, 7% noise, and a blurry observation with a moderate-size blur and 2% noise. For each degradation scenario, we randomly selected 4 images from a subset of the Berkeley Segmentation dataset [23] (roughly 700×500 pixels), and reconstructed images using a MAP estimator with

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

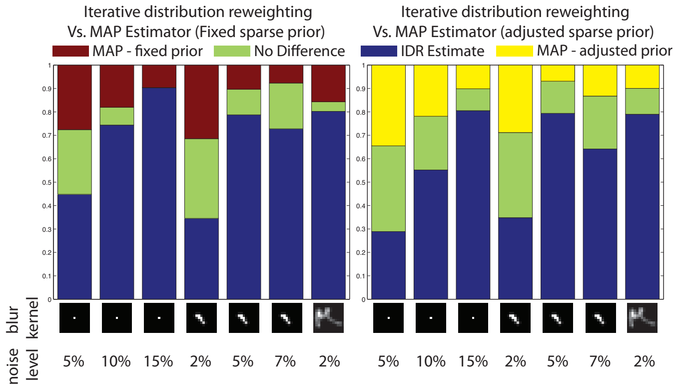


Fig. 12: *We conducted a user study to test our impression that IDR reconstructions are visually more pleasing than MAP estimates. The blue region corresponds to the fraction of users that favored IDR over MAP estimators. When the image degradation level is small, users did not show a particular preference, but as the image degradation level increases, users favored images reconstructed using IDR.*

a fixed sparse prior (i.e., the same sparse prior across the whole image), an adjusted sparse prior, and IDR.

We showed users two images side-by-side, one reconstructed using our algorithm and another reconstructed using one of the two MAP estimators (i.e., fixed or adjusted). We asked users to select an image that is more visually pleasing and give reasons for their choice. Users were also given a *"No difference."* option. We randomized the order in which we place images side by side.

We collected more than 25 user inputs for each comparison, and averaged user responses for each degradation scenario (Figure 12). When the degradation level is low (5% noise or a small blur with 2% noise), users did not prefer a particular algorithm. In such cases, the observation term is strong enough to reconstruct visually pleasing images regardless of the prior and/or the reconstruction algorithm. When the degradation level is high, however, many users clearly favored our results. User comments pointed out that realistic textures in trees, grass, and even in seemingly flat regions, such as gravel paths, are important for visual realism. Users who favored MAP estimates preferred clean renditions of flat regions and were not disturbed by piecewise smooth textures (some even found it artistic.) Individual users consistently favored either our result or MAP estimates, suggesting that image evaluation is subjective in nature.

6 Conclusion

We have developed an iterative deconvolution algorithm that matches the gradient distribution. Our algorithm bridges the energy minimization methods for deconvolution and texture synthesis. We show through a user study that matching derivative distribution improves the perceived quality of reconstructed images. The fact that a perceptually better image receives lower PSNR/SSIM suggests that there is a room for improvement in image quality assessment.

References

[1] E. P. Bennett and L. McMillan. Video enhancement using per-pixel virtual exposures. In *ACM TOG (Proc. SIGGRAPH)*, pages 845–852, New York, NY, USA, 2005. ACM.

[2] C. A. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE TIP*, 2(3):296 – 310, Mar. 1993.

[3] T. Chan and C.-K. Wong. Total variation blind deconvolution. *IEEE TIP*, 7(3):370 – 375, Mar. 1998.

[4] T. S. Cho, N. Joshi, C. L. Zitnick, S. B. Kang, R. Szeliski, and W. T. Freeman. A content-aware image prior. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[5] C. M. Christoudias, B. Georgescu, and P. Meer. Synergism in low level vision. In *IEEE ICPR*, 2002.

[6] E. Eisemann and F. Durand. Flash photography enhancement via intrinsic relighting. *ACM TOG (Proc. SIGGRAPH)*, 23:673–678, August 2004.

[7] R. Fergus, B. Singh, A. Hertzmann, S. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. *ACM TOG (Proc. SIGGRAPH)*, 2006.

[8] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE TPAMI*, 1991.

[9] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. Learning low-level vision. *IJCV*, 40(1):25 – 47, 2000.

[10] Gonzalez and Woods. *Digital image processing*. Prentice Hall, 2008.

[11] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless. Single image deblurring using motion density functions. In *ECCV*, pages 171–184, 2010.

[12] Y. HaCohen, R. Fattal, and D. Lischinski. Image upsampling via texture hallucination. In *Proceedings of the IEEE International Conference on Computational Photography (ICCP)*, 2010.

[13] D. J. Heeger and J. R. Bergen. Pyramid-based texture analysis/synthesis. In *ACM TOG (Proc. SIGGRAPH)*, 1995.

[14] N. Joshi and M. Cohen. Seeing Mt. Rainier: Lucky imaging for multi-image denoising, sharpening, and haze removal. In *Computational Photography (ICCP), 2010 IEEE Int. Conf. on*, pages 1 –8, 2010.

[15] N. Joshi, S. B. Kang, C. L. Zitnick, and R. Szeliski. Image deblurring using inertial measurement sensors. *ACM TOG (Proc. SIGGRAPH)*, 29:30:1–30:9, July 2010.

[16] N. Joshi, C. L. Zitnick, R. Szeliski, and D. Kriegman. Image deblurring and denoising using color priors. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[17] J. Kopf, C.-W. Fu, D. Cohen-Or, O. Deussen, D. Lischinski, and T.-T. Wong. Solid texture synthesis from 2D exemplars. *ACM TOG (Proc. SIGGRAPH)*, 26(3), 2007.

[18] D. Kundur and D. Hatzinakos. Blind image deconvolution revisited. *Signal Processing Magazine, IEEE*, 13(6):61 –63, Nov. 1996.

[19] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal of Optimization*, 1998.

[20] A. B. Lee, D. Mumford, and J. Huang. Occlusion models for natural images: a statistical study of a scale-invariant dead leaves model. *IJCV*, 41:35–59, 2001.

[21] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM TOG (Proc. SIGGRAPH)*, 2007.

[22] Y. Li and E. H. Adelson. Image mapping using local and global statistics. In *SPIE EI*, 2008.

[23] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.

[24] G. Matheron. *Random Sets and Integral Geometry*. John Wiley and Sons, 1975.

[25] M. Nikolova. Model distortions in Bayesian MAP reconstruction. *Inverse Problems and Imaging*, 1(2):399–422, 2007.

[26] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE TPAMI*, 12:629 – 639, 1990.

[27] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama. Digital photography with flash and no-flash image pairs. In *ACM TOG (Proc. SIGGRAPH)*, pages 664–672, New York, NY, USA, 2004. ACM.

[28] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *IJCV*, 40(1):49 – 71, Oct. 2000.

[29] R. Raskar, A. Agrawal, and J. Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. In *ACM TOG (Proc. SIGGRAPH)*, pages 795–804, New York, NY, USA, 2006.

[30] J. Rossi. Digital techniques for reducing television noise. *JSMPTE*, 87:134 – 140, 1978.

[31] S. Roth and M. Black. Fields of experts. *IJCV*, 82:205–229, 2009.

[32] S. Roth and M. J. Black. Steerable random fields. In *IEEE ICCV*, 2007.

[33] Y. Saad and M. H. Schultz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM JSSC*, 1986.

[34] U. Schmidt, Q. Gao, and S. Roth. A generative perspective on MRFs in low-level vision. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[35] E. P. Simoncelli and E. H. Adelson. Noise removal via bayesian wavelet coring. In *Proc. IEEE Int. Conf. Image Proc.*, volume 1, pages 379–382, 1996.

[36] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. Int'l Conf. Computer Vision*, pages 839–846, 1998.

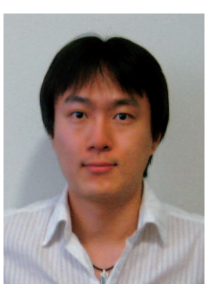
[37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 2004.

[38] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 491–498, 2010.

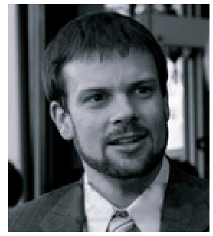
[39] O. J. Woodford, C. Rother, and V. Kolmogorov. A global perspective on MAP inference for low-level vision. In *IEEE ICCV*, 2009.

[40] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum. Image deblurring with blurred/noisy image pairs. In *ACM TOG (Proc. SIGGRAPH)*, New York, NY, USA, 2007.

[41] S. Zhu, Y. Wu, and D. Mumford. Filters, random fields and maximum entropy (frame): Towards a unified theory for texture modeling. *IJCV*, 27(2):107–126, 1998.



Taeg Sang Cho received the B.S. degree in the Department of Electrical Engineering and Computer Science from Korea Advanced Institute of Science and Technology, in 2005, and the S.M. degree and the Ph.D. degree in the Department of Electrical Engineering and Computer Science from the Massachusetts Institute of Technology, Cambridge, MA, in 2007 and 2010, respectively. He is the recipient of 2007 AMD/CICC student scholarship award, 2008 DAC/ISSCC student design contest award, 2008 IEEE CVPR best poster paper award, and 2010 IEEE CVPR outstanding reviewer award. He is a recipient of the Samsung scholarship.



C. Lawrence Zitnick received the PhD degree in robotics from Carnegie Mellon University in 2003. His thesis focused on algorithms for efficiently computing conditional probabilities in large-problem domains. Previously, his work centered on stereo vision, including cooperative and parallel algorithms, as well as developing a commercial portable 3D camera. Currently, he is a researcher at the Interactive Visual Media group at Microsoft Research. His latest work includes object recognition and computational photography. He holds over 15 patents. He is a member of the IEEE.

Neel Joshi is currently a Researcher in the Graphics Group at Microsoft Research. His work spans computer vision and computer graphics, focusing on imaging enhancement and computational photography. Neel earned an Sc.B. from Brown University, an M.S. from Stanford University, and his Ph.D. in computer science from U.C. San Diego in 2008. He has held internships at Mitsubishi Electric Research Labs, Adobe Systems, and Microsoft Research, and he was recently a visiting professor at the University of Washington. He is a member of the IEEE.

Sing Bing Kang received his Ph.D. in robotics from Carnegie Mellon University, Pittsburgh, USA in 1994. He is currently Principal Researcher at Microsoft Corporation, and his interests are image and video enhancement as well as image-based modeling. Sing Bing has co-edited two books ("Panoramic Vision" and "Emerging Topics in Computer Vision") and co-authored two books ("Image-Based Rendering" and "Image-Based Modeling of Plants and Trees"). He has served as area chair and member of technical committee for the major computer vision conferences. He has also served as papers committee member for SIGGRAPH and SIGGRAPH Asia. Sing Bing was program co-chair for ACCV07 and CVPR09, and is currently Associate Editor-in-Chief for IEEE TPAMI and IPSJ Transactions on Computer Vision and Applications.

Richard Szeliski is a Principal Researcher at Microsoft Research, where he leads the Interactive Visual Media Group. He is also an Affiliate Professor at the University of Washington, and is a Fellow of the ACM and IEEE. Dr. Szeliski pioneered the field of Bayesian methods for computer vision, as well as image-based modeling, image-based rendering, and computational photography, which lie at the intersection of computer vision and computer graphics. His most recent research on Photo Tourism and Photosynth is an exciting example of the promise of large-scale image-based rendering.

Dr. Szeliski received his Ph.D. degree in Computer Science from Carnegie Mellon University, Pittsburgh, in 1988 and joined Microsoft Research in 1995. Prior to Microsoft, he worked at Bell-Northern Research, Schlumberger Palo Alto Research, the Artificial Intelligence Center of SRI International, and the Cambridge Research Lab of Digital Equipment Corporation. He has published over 150 research papers in computer vision, computer graphics, medical imaging, neural nets, and numerical analysis, as well as the books Bayesian Modeling of Uncertainty in Low-Level Vision and Computer Vision: Algorithms and Applications. He was a Program Committee Chair for ICCV'2001 and the 1999 Vision Algorithms Workshop, served as an Associate Editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence and on the Editorial Board of the International Journal of Computer Vision, and is a Founding Editor of Foundations and Trends in Computer Graphics and Vision.

William T. Freeman is Professor of Electrical Engineering and Computer Science at MIT, working in the Computer Science and Artificial Intelligence Laboratory (CSAIL). He has been on the faculty at MIT since 2001.

He received his PhD in 1992 from the Massachussetts Institute of Technology. From 1992 - 2001 he worked at Mitsubishi Electric Research Labs (MERL), in Cambridge, MA. Prior to that, he worked at then Polaroid Corporation, and in 1987-88, was a Foreign Expert at the Taiyuan University of Technology, China.

His research interests involve machine learning applied to problems in computer vision and computational photography.